



TITLE:

非線形可積分系の応用解析の進展 :ニューロダイナミクスにおける可 積分系の視点(非線型可積分系の研 究の現状と展望)

AUTHOR(S):

中村, 佳正

CITATION:

中村, 佳正. 非線形可積分系の応用解析の進展:ニューロダイナミクスに
おける可積分系の視点(非線型可積分系の研究の現状と展望). 数理解析
研究所講究録 1993, 822: 1-13

ISSUE DATE:

1993-03

URL:

<http://hdl.handle.net/2433/83222>

RIGHT:

非線形可積分系の応用解析の進展

—ニューロダイナミクスにおける可積分系の視点—

岐阜大教育数学 中村 佳正 (Yoshimasa Nakamura)

「可積分系の応用解析」とは耳慣れない言葉だが “Applied analysis based upon the theory of nonlinear integrable systems” の意味の造語である。解説 [1, 2] で初めて用いられたもので、まだ定着した表現ではない。この数年間に「可積分系の応用解析」の対象となったテーマや問題を列記しよう。

- 1) 確率過程の線形予測問題—Riemann-Hilbert 変換法— (論文 [3])
- 2) 行列の固有値問題—スペクトル保存変形— (解説 [2])
- 3) 線形システム理論 (有理関数の空間の幾何学) (解説 [1])
- 4) 最小 2 乗法・最尤推定法 (解説 [1])
- 5) 数理計画問題—Karmarkar の力学系— (解説 [2])
- 6) 情報幾何学 (確率分布族のなす Riemann 多様体) (解説 [2])

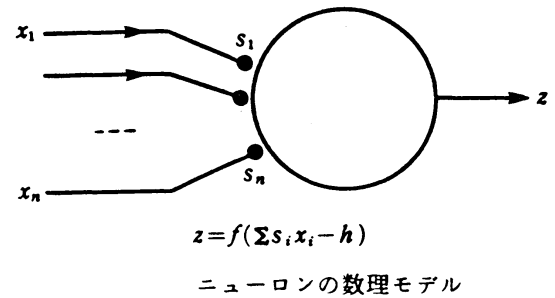
これらの文献が手に入りにくい場合は直接著者まで請求されたい。講演では、テーマ 5), 6) に加えて、「可積分系の応用解析」における最新の話題として、

- 7) ニューロダイナミクス (神経回路網の確率モデル)
- 8) 散逸力学系, 特に非線形振動論

において設定可能な可積分系の視点について報告した。1) ~ 6) と 7), 8) の違いは、前者が本来力学系とは関連の薄い問題に可積分系の切り口を与えたのに対して、後者はこれまで可積分系の視点から議論されることはなかったが、ともかくも力学系の問題である点であろう。7), 8) について、ともになんらかの意味での平均化 “averaging” を経て、ある場合、系は可積分系に移行することがわかった。言い替えれば、これまで可積分系とは程遠いとされてきた力学系においても、平均的な量を問題とすれば、可積分系によって

特徴づけられる可能性があることになる。この小論では、5)については解説[2]に譲り、6)は付録で基礎事項と結果を補足するにとどめ、7)を中心に解説する。8)については機会を改めて論じたい。

脳を形作るニューロン（神経細胞）を右図のように n 個の入力信号 x_1, \dots, x_n を受け取り、 z を出力する情報処理装置と考える [4], $0 \leq x_i \leq 1$, $0 \leq z \leq 1$. ここに、 s_1, \dots, s_n は入力信号の影響度を表すシナプス結合効率、 h はしきい値、 f は $[0, 1]$ に値をとる単調増加関数である。脳における記憶と学習を多数



のニューロンからなるネットワークのダイナミクス、具体的には、ニューロン間の結合の強さ（シナプス結合効率 s_i ）の変化と考える。ただし、自由度は一つのニューロンについて $n \approx O(10^2 \sim 4)$ 、しかも、 x_i, s_i, z は確率変数と考えるのが自然で、従来力学系の理論が対象としていたものとはかなりかけ離れている。

シナプス結合効率 s_i の変化を特徴づけるルールに Hebb の法則 (1949) がある。これはあるニューロンが興奮しているとき (z : 大)、その原因となった入力信号 x_j との結合の強さ s_j を強めていくというものである。この法則のもとで、 t を時刻、 $\gamma > 0$ を学習効率の定数として、 s_j のダイナミクスは

$$\dot{s}_j(t) = \gamma z(t) x_j(t) \quad (1)$$

と表される。これを Hebb 型学習方程式という。また、入出力関係式 $z = f(\sum s_i x_i - h)$ を簡単化して $z = \sum s_i x_i$ としたものを線形ニューロンモデルという。このときシナプス効率の変化も線形方程式に従う。

単純な Hebb 則のもとでは、時間の経過後一つのニューロンが多くの入力信号に対して反応し易くなり不合理である。そこで、ある信号に対して反応しやすくなったとき、他の信号に対する反応性を減少させる必要がある。この機構を競合学習 [4] という。具体的には von der Malsburg [5] により、 $\sum s_j = \text{一定}$ 、Oja [6] により、 $\sum s_j^2 = \text{一定}$ などが提案されている。

以上の準備のもとで、この小論で考察するのは、線形ニューロンモデルにおける記憶の

減衰項付きの Hebb 型学習方程式

$$\dot{s}_j(t) = -\gamma z^2(t)s_j(t) + \gamma z(t)x_j(t), \quad z(t) = \sum_{i=1}^n s_i(t)x_i(t) \quad (2)$$

である。通常の学習方程式，例えば [4, p.115]，では減衰項は $-s_j$ となっており，入力信号がなければシナプス結合効率 s_j はゆっくり減衰する。しかるに (2) では，出力信号 z の 2 乗に比例して減衰する。これは記憶がいったんアウトプットされればより減衰し易いことを表している。また，入出力関係は線形であるが，シナプス効率のダイナミクスは 3 次の非線形性をもつことに注意する。ベクトル $S(t) = (s_1, \dots, s_n)^T$ ， $X(t) = (x_1, \dots, x_n)^T$ を用いて (2) を

$$\dot{S}(t) = -\gamma(S^T X X^T S)S + \gamma X X^T S \quad (3)$$

と表す。先に述べたように S ， X は確率変数である。ここで，シナプス結合効率 S と入力信号 X が統計的に独立，すなわち E を可能な入力についての期待値作用素として

$$E[X^T S] = E[X^T]E[S] \quad (4)$$

を仮定する。 $E[S]$ などが存在するとして，学習方程式 (3) は

$$\dot{E}[S] = -\gamma(E[S^T]E[XX^T]E[S])E[S] + \gamma E[XX^T]E[S] \quad (5)$$

となる。これは平均学習方程式の一種である。平均学習方程式は特定の入力信号 X を使うのではなく，入力信号の分布をもとに平均化したものだから，入力信号の分布と出力信号 z を決めればシナプス結合効率の期待値 $E[S]$ だけの方程式とみなせる [4, p.118]。問題は $E[S]$ が学習によってどんな値に収束していくかを調べることである。平均化を行う前の S については，確率収束性を論じることになる。

実対称行列 $E[XX^T]$ を適当な直交行列 G で対角化し

$$C = \begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_n \end{pmatrix} = \gamma G E[XX^T] G^T \quad (6)$$

と表す。一般性を失うことなく

$$c_1 \leq c_2 \leq \dots \leq c_n \quad (7)$$

とできる。さらに,

$$R = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = GE[S] \quad (8)$$

を導入すると, (5) は

$$\dot{R} = -(R^\top CR)R + CR, \quad (9)$$

あるいは,

$$\dot{r}_j = -r_j \sum_{k=1}^n c_k r_k^2 + c_j r_j \quad (10)$$

となる。 $r_j(0) = 0$ ならば $r_j(t) = 0$ 。また,

$$\left(\sum r_j^2\right) = 2 \sum c_k r_k^2 \left(1 - \sum r_j^2\right) \quad (11)$$

だから初期値を $\sum r_j^2(0) = 1$ ととれば, 任意の t において

$$|R|^2 = \sum_{j=1}^n r_j^2(t) = 1. \quad (12)$$

これは力学系 (10) の第 1 積分である。同時に (7) を通じて $\sum E[s_j]^2 = 1$ を意味し Oja[6] による競合学習の条件 (の平均値版) に一致する。

力学系 (10) は関数

$$V(r, c) = -\frac{\sum c_k r_k^2}{2 \sum r_j^2} \quad (13)$$

についての勾配 (最急降下) 方程式 $\dot{r}_j = -\partial V / \partial r_j$ を単位球 $\sum r_j^2 = 1$ に制限したものである。従って, $V(r, c)$ は平均学習方程式 (10) の学習のポテンシャル関数 [4, p.119] とみなせる。また, 渦心点のような平衡点は存在しない。平衡点を調べるため, (13) を微分し (10) および Cauchy-Schwarz の不等式を用いると

$$\dot{V} = -\frac{\sum r_j^2 \sum c_k^2 r_k^2 - (\sum c_k r_k^2)^2}{(\sum r_j^2)^2} \leq 0. \quad (14)$$

等号は, h を定数として $r_j^2 = h c_j^2 r_j^2$, $j = 1, \dots, n$ のとき。ゆえに, もし最大固有値 c_n が重複度 1 ならば, 平衡点は

$$R = (0, \dots, 0, \pm 1)^\top, \quad h = \frac{1}{c_k^2} \quad (15)$$

と表される。

別の視点から (10) を考察しよう。行列 A, B の交換子積 $AB - BA$ を $[A, B]$ と表す。適当な行列を選ぶと $\dot{L} = [B, L]$ の形に表される非線形力学系は Lax 表示を許容するといわれる。 L, B は線形微分作用素でもよい。この表示は可積分系であるための十分条件と考えられている。平均学習方程式 (10) は以下の 2 重括弧の Lax 表示をもつ、

$$\dot{L} = \left[\frac{1}{2}C, L \right], \quad L = RR^T = \begin{pmatrix} r_1^2 & r_1 r_2 & \cdots & r_1 r_n \\ r_1 r_2 & r_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ r_1 r_n & \cdots & \cdots & r_n^2 \end{pmatrix}. \quad (16)$$

L の固有値は 0 または 1. 通常, Lax 表示からは $\text{tr}(L^k)$, $k = 1, 2, \dots, n$ として独立な第 1 積分が得られる。特に, $\text{tr}(L^2)$ は Hamilton 関数。しかしながら, この場合は $L^k = L$ だから, $\text{tr} L = \sum r_j^2 = 1$ 以外の第 1 積分は得られない。Hamilton 関数の存在もこのままではよくわからない。そこで, 以下 (10) を拘束条件付きの Hamilton 系とみなせることを明らかにしよう。 $r_j(0) = 0$ ならば $r_j(t) = 0$ なので, ここでは初期値 $\{r_j(0)\}$ を \mathbb{R}^n の単体 $\{x_j | \sum_{j=1}^n x_j = 1, x_j \geq 0\}$ の内部にとり, ただし,

$$x_j(t) = r_j^2(t). \quad (17)$$

x_j に加えて c_j も力学変数とみなし, (10) を書き直して $\mathbb{R}^n \oplus \mathbb{R}_+^n$ 上の力学系

$$\dot{x}_j = -2x_j \sum_{k=1}^n c_k x_k + 2c_j x_j, \quad \dot{c}_j = 0 \quad (18)$$

を考える。ここに, $\mathbb{R}_+ = \{y \in \mathbb{R} | y > 0\}$ 。

次に, $\mathbb{R}^n \oplus \mathbb{R}_+^n$ 上にシンプレクティック 2 形式

$$\omega = - \sum_{i=1}^n \frac{dc_i \wedge dx_i}{x_i} \quad (19)$$

を導入する。付随する Poisson 括弧は

$$\{f, g\} = \sum_{i=1}^n x_i \left(\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial c_i} - \frac{\partial g}{\partial x_i} \frac{\partial f}{\partial c_i} \right). \quad (20)$$

$\mathbb{R}^n \oplus \mathbb{R}_+^n$ 上の Hamilton 関数 $H_0 = \sum c_k^2$ を考える。運動方程式は $\dot{c}_j = \{c_j, H_0\} = 0$, $\dot{x}_j = \{x_j, H_0\} = 2x_j c_j$ 。これより, $\ddot{x}_i = 4c_i^2 x_i$ (斥力の調和振動子)。 c_0 を定数として $\phi_1 = \sum x_i - 1$, $\phi_2 = \sum c_i - c_0$ とおく。このとき, $\{\phi_1, \phi_2\} = \phi_1 + 1$, $\{\phi_1, H_0\} = 2 \sum c_i x_i$,

$\{\phi_2, H_0\} = 0$. Hamilton 関数 H_0 を $2n - 2$ 次元部分空間 $M = \{(c, x) \in \mathbb{R}^n \oplus \mathbb{R}_+^n \mid \phi_1 = 0, \phi_2 = 0\}$ に制限して, $H_0 = H + \alpha_1 \phi_1 + \alpha_2 \phi_2$ とおく. ここに, α_j は $\{\phi_j, H\} = 0$ なるように定める. 結果は, $\alpha_1 = 0, \alpha_2 = 2 \sum c_i x_i$. ゆえに,

$$H = \sum_{k=1}^n c_k^2 - 2 \sum_{j=1}^n c_j x_j \left(\sum_{k=1}^n c_k - c_0 \right) \quad (21)$$

を Hamilton 関数とする運動方程式は第 1 積分をもち自由度 $n - 1$ である. 運動方程式を書き下すと

$$\dot{x}_j = \{x_j, H\} = -2x_j \sum_{k=1}^n c_k x_k + 2c_j x_j - 2x_j^2 \left(\sum_{k=1}^n c_k - c_0 \right), \quad (22)$$

$$\dot{c}_j = \{c_j, H\} = -2c_j x_j \left(\sum_{k=1}^n c_k - c_0 \right) \quad (23)$$

となり, $\sum c_j(0) = c_0, \sum x_j(0) = 1$ ととれば (18) に帰着する.

以上をまとめると, ニューロダイナミクスの平均学習方程式 (10) は単体の頂点を平衡点とする勾配系であると同時に斥力の調和振動子を偶数次元部分空間に制限するという意味で拘束づけられた可積分な Hamilton 系であると結論される.

Lax 表示をもつ可積分な勾配系の例としては, この他解説 [2] で扱っている Karmarkar の力学系や確率分布空間上の勾配系 (43) があるが, このクラスの最も基本的な力学系はやはり Moser[7] が考察した有限非周期戸田方程式であろう. しかも, この戸田方程式は平均学習方程式 (10) の特殊化と位置づけられる. 以下このことを述べよう.

Moser による戸田方程式は Flaschka の変数を用いて

$$\dot{a}_k = a_k(b_{k+1} - b_k), \quad \dot{b}_k = 2(a_k^2 - a_{k-1}^2), \quad (24)$$

$k = 1, \dots, n$ と表される. ただし, 物理的な要請から

$$a_0 = 0, \quad a_n = 0, \quad a_j(t) > 0 \quad \text{for } j = 1, \dots, n-1 \quad (25)$$

なる付帯条件が課せられている. a_k, b_k は正準変数ではないが, これを使うと戸田格子の Hamilton 関数は

$$H_{\text{TL}} = 4 \sum_{k=1}^{n-1} a_k^2 + 2 \sum_{k=1}^n b_k^2 \quad (26)$$

と書かれる。また、運動方程式 (24) の Lax 表示として

$$\dot{M} = [M_U - M_U^T, M], \quad M = \begin{pmatrix} b_1 & a_1 & & 0 \\ a_1 & b_2 & \ddots & \\ & \ddots & \ddots & a_{n-1} \\ 0 & & a_{n-1} & b_n \end{pmatrix} \quad (27)$$

がよく知られている。ここに、 M_U は M の強上三角部をとるの意。Moser[7] は (24) の作用角変数を求める過程で λ についての有理関数

$$f(\lambda; t) = e_n^T (\lambda I - M)^{-1} e_n, \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (28)$$

を導入した。 M において条件 (25) より $f(\lambda; t)$ は

$$f(\lambda; t) = \sum_{j=1}^n \frac{r_j^2(t)}{\lambda - \lambda_j}, \quad \lambda_1 > \lambda_2 > \cdots > \lambda_n, \quad \sum_{j=1}^n r_j^2 = 1 \quad (29)$$

と一意に部分分数展開される。 M の固有値 λ_j について等号がないことに注意する。もちろん、 λ_j は t に依らない。 $f(\lambda; t)$ を t で微分して (24) を用いると

$$\dot{r}_j = r_j \sum_{k=1}^n \lambda_k r_k^2 - \lambda_j r_j \quad (30)$$

を得る。これは $\lambda_j \rightarrow -c_j$ とすれば、 $c_i \neq c_j$ である場合の (10) に一致する、cf. (7)。Moser の方程式 (30) の勾配系としての側面や拘束づけられた Hamilton 系の構造はそれぞれ [7], [8] で論じられている。また、 t の指数関数を用いて求積できることもわかっている [7]。

まとめ この小論では、線形ニューロンモデルにおける記憶の減衰項付きの Hebb 型学習方程式 (2) が、入力信号についての平均化を行うことで、Moser による戸田方程式を一般化した可積分な勾配系 (10) となることを明らかにした。さらに、この平均学習方程式の平衡点や Hamilton 構造などを考察した。特に、第 1 積分の存在はニューロンモデルに対する競合学習の条件となっていることを示した。

なお、この研究の一部は文部省科研費 03804005, 04804005 の援助を受けている。

付録 — 情報幾何学における双対平坦構造と可積分系 —

「情報幾何学」とは情報のなす集合を連続的な位相をもつ空間とみなし、微分幾何学的考察を通じてその本質をとらえようという試みである。統計的推論における高次の漸近展開や線形・非線形システムの空間への応用における成功（小原氏の講演参照）に加え、双対 affine 接続という微分幾何学として新しい概念をもたらすなど興味深い対象と考えられる。成書として主唱者の甘利氏による [9] がある。最近, [10] において, 空間の双対平坦性に起因してパラメータの最尤推定値に収束する可積分な勾配系が存在することが示された。ここでは情報幾何学を「確率分布族のパラメータ空間の微分幾何学」として少し狭い意味でとらえ, [9] に従って特にその双対平坦構造について解説する。

ある確率分布の密度関数を $p(x, \theta)$ とする。確率変数 x は連続・離散・多変数を問わない。 $\theta = (\theta^1, \dots, \theta^k)$ は平均, 標準偏差など個々の分布を特定する k 次元パラメーターである。考察する確率分布の全体を $S = \{\theta | p(x, \theta)\}$ とかく。簡単な例として, (x について 1 次元の) Gauss 分布をとりあげよう。密度関数は

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R} \quad (31)$$

である。もし, θ を $\theta = (\mu, \sigma)$ ととれば, 2 次元空間 S は上半面 $S = \{(\mu, \sigma) | \mu \in \mathbb{R}, \sigma > 0\}$ となる。 S の分布間の (統計的意味での) 分離度を計る際, たとえ σ が同一で μ についての差が同じであっても σ の値いかんで分離度が変化する。従って S における距離は Euclid 計量では都合が悪く個々の確率分布に内在する構造を反映したものとして導入されねばならないことがわかる。 $p(x, \theta)$ に基づいて期待値をとる操作を $E[\]$ と表し, x の空間の適当な support 上で $p(x, \theta) > 0$ などの仮定のもとで, 対数尤度関数 $\ell = \log p(x, \theta)$ を導入する。このとき,

$$\langle \partial_i, \partial_j \rangle = E[\partial_i \ell \cdot \partial_j \ell] \quad (32)$$

は Fisher 情報量と呼ばれる, ただし, $\partial_i = \partial/\partial\theta^i$. Rao(1945) は S の計量として $g_{ij}(\theta) = \langle \partial_i, \partial_j \rangle$ を採用することを提案し, Chentsov(1972) はこの計量 (Fisher 計量) が x のスケール変換 $x \rightarrow x' = h(x)$ のもとで不変な唯一の計量であることを示した。今日では, 多くの確率分布についてその族のなすパラメーター空間 S が $g_{ij}(\theta)$ を Riemann 計量とする k 次元の Riemann 多様体となることが知られている。Gauss 分布の例では θ の適当な選択により $g_{ij}(\theta)$ は Poincaré 計量と表せ, その結果 S は擬球とみなされる。

次に, Riemann 多様体 (S, g) について Levi-Civita 接続

$$\Gamma_{ijk} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) \quad (33)$$

を準備する. θ を固定したとき $\partial_i \ell$ の線形独立性を仮定する. $\partial_i \ell(\theta)$ は θ における接空間 T_θ の基底をなし, E と ∂_i が可換であれば $E[\partial_i \ell] = 0$ を満たす. 平行移動により $\partial_i \ell(\theta + d\theta)$ が T_θ に移されるためには, 接続を Levi-Civita 接続に限る必要はなく, α をパラメータとして

$$\Gamma_{ijk}^{(\alpha)} = \Gamma_{ijk} - \frac{\alpha}{2} T_{ijk}, \quad T_{ijk} = E[\partial_i \ell \cdot \partial_j \ell \cdot \partial_k \ell] \quad (34)$$

であればよい. この affine 接続の 1-パラメータ族を α -接続という. $\alpha = 0$ に相当するのが Levi-Civita 接続. $\alpha = \pm 1$ の場合は, 情報幾何学では特別な意味をもち, それぞれ, ∇ -接続, ∇^* -接続と呼ばれている. また, α -接続に関して定まるねじれ率テンソルは任意の α についてゼロである. さらに, Riemann-Christoffel の α -曲率テンソル

$$R_{ijkm}^{(\alpha)} = (\partial_i \Gamma_{jk}^{(\alpha)n} - \partial_j \Gamma_{ik}^{(\alpha)n}) g_{nm} + (\Gamma_{inm}^{(\alpha)} \Gamma_{jk}^{(\alpha)n} - \Gamma_{jnm}^{(\alpha)} \Gamma_{ik}^{(\alpha)n}) \quad (35)$$

を導入する. Einstein の総和則を用いている. S の任意の点で $R_{ijkm}^{(\alpha)}(\theta) = 0$ であるとき S は α -平坦という. このとき $\Gamma_{jkm}^{(\alpha)}(\theta) = 0$ なる α -affine 座標系 θ が存在する. Gauss 分布では曲率テンソルはゼロもしくは $R_{1212}^{(\alpha)} = (1 - \alpha^2)/\sigma^4$ のみによって表すことができ, スカラー曲率は $\kappa = -(1 - \alpha^2)/2$ となる. 明らかに $\alpha = \pm 1$ 平坦. $\theta = (\mu/\sigma^2, -1/(2\sigma^2))$, $\theta' = (\mu, \mu^2 + \sigma^2)$ は, それぞれ, 1-affine, -1-affine 座標系である. 以下, -1-affine 座標系を $\eta = (\eta_1, \dots, \eta_k)$ とかき, $\partial^j = \partial/\partial \eta_j$ と約束する. また, $\langle \partial_i, \partial^j \rangle = \delta_i^j$ であるとき, 座標系の組 (θ, η) は双対的という. このとき, $\partial_i = g_{ij} \partial^j$. 双対座標系の存在と α -平坦性は深く結びついている. [9] よりいくつかの基本定理を引用する.

定理 1 Riemann 多様体 (S, g) が affine 接続 ∇ , ∇^* に関しともに平坦でねじれ率テンソルがゼロならば, それぞれ, ± 1 -affine であるような双対座標系 (θ, η) が存在する.

多くの確率分布族のなす Riemann 多様体が $\alpha = \pm 1$ で双対平坦であることがわかっている. 双対平坦性は情報幾何学の核心部分と考えられる. 実際,

定理 2 Riemann 多様体 (S, g) が双対座標系 (θ, η) をもつならば,

(i) 関数 $\psi(\theta)$, $\phi(\eta)$ で関係式

$$\psi(\theta) + \phi(\eta) - \theta^i \eta_i = 0 \quad (36)$$

を満たすものが存在する, ただし, $\theta^i \eta_i = \sum_{i=1}^k \theta^i \eta_i$.

(ii) 座標系 (θ, η) で定まる点において Fisher 計量は

$$g_{ij}(\theta) = \partial_i \partial_j \psi(\theta), \quad g^{ij}(\eta) = \partial^i \partial^j \phi(\eta) \quad (37)$$

と表される.

この結果, 各座標系は $\theta^i = \partial^i \phi(\eta)$, $\eta_i = \partial_i \psi(\theta)$ と書かれる. (36) の定める変換 $\theta \leftrightarrow \eta$ は一種の Legendre 変換である. よく知られているように Legendre 変換は, 古典力学では Lagrange 形式から Hamilton 形式への書き換え, 熱力学では状態変数の変換で用いられる操作である. また, $\psi(\theta)$ は分布の $(-1) \times$ エントロピーとみなせる.

ここで 2 つの分布間の「距離」について論じる. 双対平坦な Riemann 多様体 S の 2 点を $P(\theta_P, \eta_P)$, $Q(\theta_Q, \eta_Q)$ とする. P から Q への ∇ -divergence を

$$D(P, Q) = \psi(\theta_P) + \phi(\eta_Q) - \theta_P^i \eta_{Qi} \quad (38)$$

と定義する. ただちに, $D(P, Q) = \psi(\theta_P) - \psi(\theta_Q) - (\theta_P^i - \theta_Q^i) \eta_{Qi}$ が従う. さらに, $D(P, Q) \geq 0$, 等号は $P = Q$ に限られる. しかし, $D(P, Q) = D(Q, P)$ や 3 角不等式は一般には成り立たないので $D(P, Q)$ は正確な意味での距離関数とはいえない. 定理 2 より $\partial^2 D(P, Q) / \partial \theta_P^i \partial \theta_P^j = g_{ij}(\theta_P)$ がわかる. P と Q が近く $\theta_Q = \theta_P + d\theta_P$ とかけるとき

$$D(P, Q) = \frac{1}{2} g_{ij}(\theta_P) d\theta_P^i d\theta_P^j + o(|d\theta_P|^3) \quad (39)$$

と表され, Fisher 計量による分離度と関係づけられる. ∇ -divergence は距離ではないものの, 一種の Pythagoras の定理や射影定理が成り立つ.

定理 3 ± 1 -平坦な曲線を, それぞれ, ∇ -測地線, ∇^* -測地線という. 3 点 P, Q, R について P と Q を結ぶ ∇ -測地線が Q と R を結ぶ ∇^* -測地線と点 Q で直交するとき

$$D(P, Q) + D(Q, R) = D(P, R). \quad (40)$$

S の部分集合 W が -1 -convex であるとは任意の 2 点 $P_1, P_2 \in W$ について P_1 と P_2 を結ぶ ∇ -測地線が一意に存在することをいう.

定理 4 S の閉部分集合 W が -1 -convex で滑らかな境界 ∂W をもつとする. このとき, W の外点 P の W における ∇ -divergence の意味での最良近似は, P を ∇^* -測地線により ∂W に射影した点である.

さらに、一方の仮説の確からしさを表す $p_1(x, \theta_1)$ を他方の仮説の確かさを表す $p_2(x, \theta_2)$ から区別するための尺度（分離度）として仮説検定で用いられる Kullback-Leibler divergence を準備する。これは $p_1(x, \theta_1)$, $p_2(x, \theta_2)$ に対応する S の点を、それぞれ、 P_1, P_2 とするとき、

$$K(P_1, P_2) = E[\log(p_1(x, \theta_1)/p_2(x, \theta_2))] \quad (41)$$

と書かれる、ただし、期待値は $p_1(x, \theta_1)$ についてとるものとする。密度関数が確率変数 $x = x_i$ とパラメータ θ の適当な選択の結果

$$p(x, \theta) = \exp(\theta^i x_i - \psi(\theta)) \quad (42)$$

と表される分布を指数関数族 (exponential family) に属するという。Gauss 分布を含む非常に多くの確率分布がこのクラスに入るが、指数分布族については Kullback-Leibler divergence $K(P_1, P_2)$ は ∇ -divergence $D(P_2, P_1)$ に一致する。

さて、双対平坦構造をもつ確率分布族のなす Riemann 多様体 (S, g, ∇, ∇^*) を議論の舞台としよう。ポテンシャル $\psi(\theta)$ に対する S 上の勾配系

$$\frac{d\theta^i}{dt} = -g^{ij}(\theta) \partial_j \psi(\theta) \quad (43)$$

を考える。講演で述べたように以下の事実が明らかになっている [10]。

I. 勾配系は θ の双対座標系 $\eta = (\eta_1, \dots, \eta_k)$ により常に線形化可能、ゆえに可積分。流れはエントロピー $\psi(\theta)$ の最大点に指数関数的に収束する。

II. 特に、 S が偶数次元、すなわち $k = 2m$ であれば、勾配系は完全積分可能な Hamilton 系をなし、その等エネルギー曲面は Euclid 空間 \mathbb{R}^m に微分同相である。奇数次元の場合も本質的には適当な射影により偶数次元の場合に帰着できる。

なお、 $\psi(\theta)$ の代りに Kullback-Leibler divergence を用いれば、指数分布族の場合常に線形化可能で、しかも統計学的意味が明確となる。この点については、藤原氏の講演で詳しく述べられた。

情報幾何学において設営が可能にみえる可積分系の視点の前進基地としてはこのほか次のような地点が有望であろう。

(i) 同じく、凸領域や凸関数が現れ、Legendre 変換により線形化可能な力学系のクラスとして、Dikin や Karmarkar 等による線形計画問題の内点流がある。ある場合には Lax

対表示が報告されている [2]. ある凸関数の勾配方向に交互に双対 affine 座標系をたどることで内点アルゴリズムの大域的収束性を実現している点など舞台設定は情報幾何学と共通している.

(ii) $\psi(\theta)$ は Kähler ポテンシャルの類似であるから, hyper Kähler 計量をその特殊解にもつ自己双対 Einstein 方程式を通じて非線形可積分系の理論とかかわると予想される. 実際, 多項分布の場合, 双対ポテンシャル $\phi(\eta)$ は Fubini-Study 計量に他ならない [10]. また, Hitchin 等 (1987) により, ある種の Kähler ポテンシャルと Legendre 変換を用いて多重 Eguchi-Hanson 解, Taub-NUT 型の自己双対インスタントン解や Calabi 計量を含む hyper Kähler 計量が構成されている.

参考文献

- [1] 中村佳正, 非線形可積分系の応用解析の試み — 力学的対称性による非線形問題へのアプローチ, 計測と制御 31(1992), 872-877.
- [2] 中村佳正, 非線形可積分系の応用解析の展開, 応用数理 2(1992), 330-342.
- [3] Nakamura, Y., A new approach to the linear prediction problem for infinite stationary processes in terms of Riemann-Hilbert transformation theory, IMA J. Math. Control & Inform., 6(1989), 347-358.
- [4] 甘利俊一, 神経回路網モデルとコネクショニズム, 東大出版, 1989.
- [5] von der Malsburg, C., Self-organization of orientation sensitive cells in the striate cortex, Kybernetik, 14(1973), 85-100.
- [6] Oja, E., A simplified newron model as a principal component analyzer, J. Math. Biology, 15(1982), 267-273.
- [7] Moser, J., Finitely many mass points on the line under the influence of an exponential potential - An integrable system, in: Dynamical Systems, Theory and Applications, (ed. J. Moser), Springer-Verlag, 1975, pp. 467-497.
- [8] Deift, P., Lund, F., and Trubowitz, E., Nonlinear wave equations and constrained harmonic motion, Commun. Math. Phys. 74(1980), 141-188.

- [9] Amari, S., Differential-Geometric Methods in Statistics, Lec. Notes in Statist. Vol. 28, Springer-Verlag, Berlin, 1985.
- [10] Nakamura, Y., Completely integrable gradient systems on the manifolds of Gaussian and multinomial distributions, to be published in Japan J. Indust. Appl. Math., 1993.